VIReC RESEARCHERS' GUIDE TO VA DATA

Data Issues Brief March 2004

## 1. FY2003 & FY2004 Medical SAS Datasets Missing Race& Ethnicity Data

Researchers utilizing VHA Medical SAS Datasets for FY2003 and FY2004 are facing a lack of data on patients' race and ethnicity. Information received March 2, 2004, by VIReC and others from National Data Systems indicated that this resulted from implementation of new race and ethnicity variables (see *Data Issues Brief*, June 2003). Specifically, the NDS stated, "What you are seeing in FY03 with missing values is the manifestation of the old value no longer being transmitted to the national database and the new values not yet populated." NDS response currently focuses on improving the collection of information to populate the new variables.

## 2. VIReC Exploring Options For Researchers Needing Data on Race

VIReC recognizes the importance of these data for researchers and is pursuing alternative approaches to resolution of the missing data problem. In the meantime, it is suggested that researchers mine the Medical SAS Datasets for fiscal years prior to 2003 to obtain a value for race to the extent possible.

The validity and reliability of the VHA national data on race has not been definitively assessed in the published literature. To provide some understanding of the quality of the VA race data, VIReC compared the value of the race variable for veterans from the FY1999 Inpatient Main and Outpatient Visit Medical SAS Datasets to the race variable in the Medicare Denominator Files CY1999 – CY2001 (see the VIReC Web site for details on the VA-Medicare data). The Medical SAS Datasets were merged to the Denominator Files by SSN, date of birth and gender. Because some racial and ethnic code values in Medicare data ("Asian", "Hispanic", etc.) have been shown in the literature to be less accurate than the "White" and "Black" code values, we re-coded race data in the VA and Medicare files to two values: (1) "Black," indicating that the veteran had at least one occurrence in the years studied of race coded as "Black" or "Hispanic Black," and (2) "Non-Black," indicating that the veteran had no occurrence in the years studied of race coded as "Black" or "Hispanic Black". (A very small number of veterans had variable codes for race in either the VA or Medicare data sources).

Of the 3,370,115 veterans selected from the FY1999 Medical SAS Datasets, 1,679,696 were found in the Medicare files; of these, 1,104,508 had a race value coded in both the VA and Medicare files. Using the Medicare data as the gold standard, the individual-level data on race for these veterans, coded in the new "Black/Non-Black" values, were compared between the two sources. Preliminary results follow:

**Comparison of the Race Codes in the Medical SAS Datasets and the Medicare Denominator File**

| VA Data | Medicare Data | | |
|---|---|---|---|
| | Black | Non-Black | Total |
| Black | 136,918 | 10,057 | 146,975 |
| Non-Black | 10,452 | 947,081 | 957,533 |
| Total | 147,370 | 957,138 | 1,104,508 |

| Agreement Measures | | | | |
|---|---|---|---|---|
| Sensitivity | Specificity | PPV | NPV | Kappa |
| 92.9 | 98.9 | 93.2 | 98.9 | 92.0 |

Sensitivity measures the percentage of "Black" veterans who are correctly classified with a race code of "Black". The results show that 92.9% of the veterans in our subset from the Medical SAS Datasets with a race code populated as "Black" are correctly classified using the Medicare data as the gold standard. Specificity measures the percentage of individuals who should not be classified as "Black" who have a race code of "Non-Black". In this comparison, the specificity is very high, 98.9%, indicating that the probability of misclassifying an individual as "Black" is only about 1%. The results indicate that the race value from previous years Medical SAS Datasets may provide a fairly accurate source for race codes.

Researchers utilizing previous years' VA data on race face two major limitations. First, over 30% of the race values in the VA data are missing, and, second, race values for veterans enrolled since the beginning of FY2003 probably will not be found. Alternative sources will be needed. In our comparison we found race codes in the Medicare data for almost all of the 573,352 veterans with no VA race code value. VIReC plans to investigate the availability of alternative sources.

### 3. New Office of Research & Development Contact for Real SSN Access

Researchers' requests for access to real Social Security Number information in the corporate databases maintained at the Austin Automation Center (AAC) require review and approval by the Office of Research and Development (ORD) before access is granted through the Automated Computer Resources System (ACRS). Beginning March 1, 2004, ACRS staff at VHA facilities will route requests for research use of real SSN data to Brenda Cuccherini, Ph.D., M.P.H. See the VIReC Web site for details on requesting access to VA maintained at the AAC.

### 4. Financial and Clinical Data Marts (FCDM) Open For Business

The FCDM is now providing DSS National Data Extracts (NDE), including Fact of Lab (LAB), in Web-accessible online analytical processing (OLAP) cubes. FCDM's ProClarity Analytics® Web server provides online help with navigating the site and manipulating the data display. First-time users of the FCDM can access it through the VA Intranet at the VISN Support Service Center's KLF menu site (http://klfmenu.med.va.gov/dss_reports). A VA username and password are required.

### 5. VIReC Beta Testing Redesigned Web Site

On March 1, 2004, VIReC began beta testing of its redesigned Web site. In addition to design and navigation changes intended to provide more accessible and easily found information for researchers using VA data, five sections of the site have been rewritten. These are pages entitled Toolkit for New Users of VA Data, Pharmacy Benefits Management (PBM) Database, VHA Decision Support System (DSS), Veterans Health Information Systems and Technology Architecture (VistA), and the VA-Medicare Data For Research. Comments and suggestions for the site are welcomed. The Web site is at www.virec.research.med.va.gov.